



Utilisation du logiciel R pour l'identification de nouvelles cibles et régulateurs du protéasome

Céline Pellentz, Cosmin Saveanu, Alain Jacquier, Anne Peyroche

► To cite this version:

Céline Pellentz, Cosmin Saveanu, Alain Jacquier, Anne Peyroche. Utilisation du logiciel R pour l'identification de nouvelles cibles et régulateurs du protéasome. 1ères Rencontres R, Jul 2012, Bordeaux, France. hal-00717558

HAL Id: hal-00717558

<https://hal.science/hal-00717558>

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Utilisation du logiciel R pour l'identification de nouvelles cibles et régulateurs du protéasome

C. Pellentz^a, C. Saveanu^b, A. Jacquier^b et A. Peyroche^a

^a CEA Saclay
IBITECS, SBIGEM, LMARGE
F-91191, France
celine.pellentz@cea.fr
anne.peyroche@cea.fr

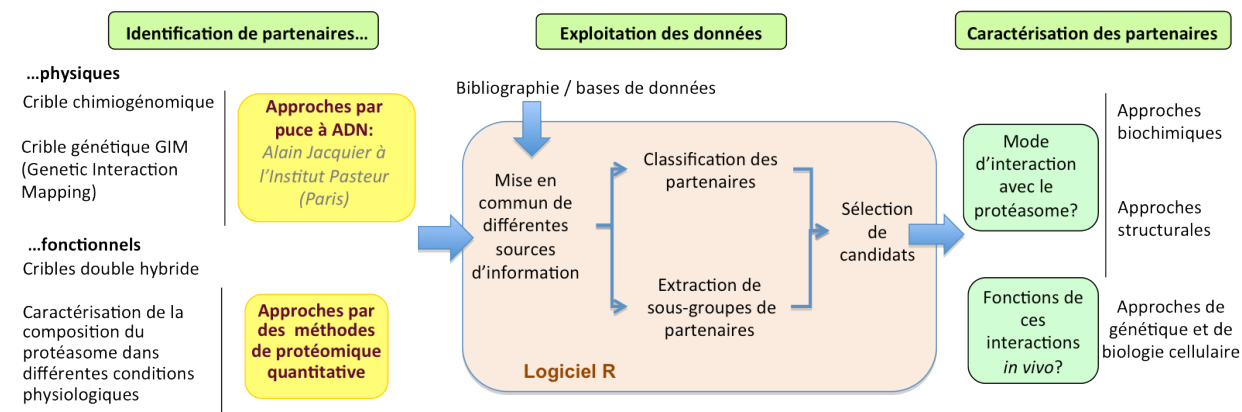
^b Institut Pasteur
Génétique des Interactions Macromoléculaires
25-28, Rue du Docteur Roux
75724 Paris Cedex 15 France
cosmin.saveanu@pasteur.fr
alain.jacquier@pasteur.fr

Mots clefs : Biologie, Analyse de cribles et de données, Hiérarchisation, Clusterisation.

Le protéasome est une protéase multimérique composée de différents sous-complexes obtenus par l'assemblage ordonné de dizaines de polypeptides. Il est présent chez toutes les cellules eucaryotes où il constitue l'unité catalytique du système Ubiquitine-protéasome (UPS). Son activité de dégradation des protéines, à la fois cytoplasmique et nucléaire est essentielle à de nombreux processus cellulaires [1]. Le protéasome est une machinerie plastique et dynamique qui est elle-même soumise à régulation [2]. Des dysfonctionnements de l'UPS participent à la pathogenèse de maladies telles que les cancers ou les maladies neurodégénératives. Il semble donc important d'identifier i) les processus cellulaires dans lesquels le protéasome est impliqué, ii) les facteurs influençant son activité.

Dans ce but, nous développons un projet pour identifier et caractériser les partenaires physiques et fonctionnels du protéasome par une approche multi-technique chez la levure *S. cerevisiae*. Nous déterminerons ensuite s'ils sont fonctionnellement conservés chez les Mammifères.

Ce projet peut être divisé en trois étapes :



1/ L'identification de partenaires

La première étape consiste à identifier des partenaires physiques et fonctionnels du protéasome par la réalisation de cribles à grande échelle utilisant le modèle cellulaire *Saccharomyces cerevisiae*.

2/L'exploitation des données grâce au logiciel R

L'utilisation du logiciel R permet d'exploiter les résultats des cribles et de croiser ces données avec d'autres données d'interactions issues de résultats du laboratoire ou de ressources bibliographiques.

Les cribles utilisés pour identifier les partenaires fonctionnels du protéasome reposent sur la technologie des puces à ADN. La quantité des données quantitatives qui en sont issues doit être gérée et exploitée pour contrôler la reproductibilité des cribles et distinguer les répondants du bruit de fond. Les résultats de ses cribles peuvent ensuite être hiérarchisés pour essayer d'identifier des groupes répondants de façon similaire ou opposé.

Le deuxième aspect de l'utilisation du logiciel R dans ce projet est le regroupement de données expérimentales d'origines variées mais ayant pour point commun l'étude des partenaires du protéasome. L'intérêt de cette approche est double : i) chaque technique expérimentale présente des biais et bruits de fond différents, comparer des résultats de techniques très différentes permet de s'affranchir du bruit de fond de chacune et d'être plus sensible ; ii) les techniques à grande échelle, qui sont de plus en plus utilisées, génèrent une très grande quantité de données qu'il est impossible de mémoriser et comparer aux données existantes antérieures sans outil informatique, compiler ces données grâce au logiciel R permet de les pérenniser et de pouvoir donner du poids à des données qui ont peu de poids isolées mais qui deviennent significatives quand elles sont retrouvées dans d'autres jeux de données.

Enfin, la classification des interactants potentiels identifiés suivant différents critères permet de faire ressortir des candidats potentiels non identifiés auparavant. Classer les interactants selon le critère du nombre de techniques différentes l'ayant mis en évidence permet par exemple de s'affranchir du bruit de fond de chaque technique. On peut ainsi identifier de nouveaux acteurs potentiels du fonctionnement du protéasome et/ou mettre en évidence de nouvelles connexions entre le système UPS et d'autres voies cellulaires.

La banque d'interactants ainsi générée peut ensuite être interrogée pour identifier des interactants potentiels du protéasome dans son ensemble ou des interactants spécifiques d'une sous-unité donnée (parmi la trentaine de sous-unités au total) du protéasome.

3/Caractérisation de candidats potentiels

Enfin, la dernière étape consiste en la caractérisation de ces candidats potentiels, en combinant des approches biochimiques et génétiques classiques.

Cette démarche originale pourrait permettre d'identifier des partenaires du protéasome non encore identifiés grâce à la puissance de la comparaison de données issues de techniques très différentes qui permet de distinguer le bruit lié à chaque technique des interactants probablement significatifs.

Références

- [1] Bedford L., Paine S., Sheppard P.W., Mayer R.J., Roelofs J. (2010) Assembly, structure and function of the 26S proteasome. Trends in Cell Biology 20:391-401
- [2] Glickman M.H. and Raveh D., (2005) Proteasome plasticity. FEBS Lett, 579(15): p3214-23